Project Closure Document

End-To-End Solution using Globus Online Integrated with glideinWMS

1. Table of Contents

2. *Approvals*

glideinWMS Project Representative:	Signature:		Date:
	Print Name:	Burt Holzman	
	Title:		

Intensity Frontier Representative:	Signature:		Date:
	Print Name:	Lee Lueking	
	Title:		

REX Operations Representative:	Signature:		Date:
	Print Name:	Margaret Votava	
	Title:		

FermiGrid/FermiCloud Representative:	Signature:		Date:
	Print Name:	Steve Timm	
	Title:		

Computing Division Representative:	Signature:		Date:
	Print Name:	Gabriele Garzoglio	
	Title:		

Project Leader & CEDPS Representative:	Signature:		Date:
	Print Name:	Parag Mhashilkar	
	Title:	Application Developer and System Analyst	

3. Document Change Log

Version	Date	Change Description	Prepared By
V 1.0	11/04/2011	First Version of the Document	Parag Mhashilkar
V 1.1	11/18/2011	Incorporated the feedback	Parag Mhashilkar

4. Project Proposal Lead

Project Leader : Parag Mhashilkar
Department : Computing Division
Group : SCF/GRID/DOCS

5. Project Documentation

The project web home page:
https://cdcvs.fnal.gov/redmine/projects/cedps-glideinwms

Following relevant documents can be found in the Fermilab's docdb -
- CEDPS is covered in tactical plan document : CD-doc-3826
- FermiCloud is covered in tactical plan document : CD-doc-4401
- FermiCloud economic model : CD-doc-4399
- glideinWMS is covered in tactical plan document : CD-doc-4394
- End-To-end Solution, project definition document : CD-doc-4406
- End-To-end Solution, stress tests using Globus Online : CD-doc-4474

6. Reason for Closing the Project

This project was an investigative effort aimed at providing a working proto-type that can retain the ownership of the output files produced by the user submitting the grid job. This system is designed to provide an End-To-End solution by integrating with the current grid processing infrastructure and the tools that are currently in use for grid submission.

The project has achieved the initial goals stated in the charter (see "Project Deliverables").

As of now there are no outstanding user requests known to the project. The gridftp service has been deployed on the FermiCloud virtual machine if-gridftp.fnal.gov. If there are any user requests in future and if the need be to add features to existing services, we propose to open a new phase of the project with a new project definition that is adequate to the changed needs of the community.

7. Project Deliverables and Milestones

The high level milestones and deliverables for the project are –

Milestones/Deliverables	Actual Deliverables and Impact
Develop tools to generate authentication and authorization list for accessing gridftp server	Tools developed as part of the sub-project http://cdcvs.fnal.gov/redmine/projects/if-gridftp-setup
Setup a test bed for testing gridftp server	Setup test VM if-gridftp.fnal.gov on FermiCloud with the if-gridftp-setup tools
Develop condor-globusonline-transfer-plugin	Plugin and scripts developed as part of sub-project https://cdcvs.fnal.gov/redmine/projects/go-condor-plugin
Integrate condor-globusonline-transfer-plugin with glideinWMS	
Deploy glideinWMS for end-to-end testing and commissioning	Setup test VM on FermiCloud, configured with plugin scripts to run test jobs. Test jobs were used to store files to the experiment disks using globus-url-copy and Globus Online.
Deploy gridftp servers the required IF experiment(s)	
Register the gridftp servers with Globus Online Service	
Close the Project	

8. Project Schedule

Following table describes the project schedule –

Milestones/Deliverables	Stakeholder	Planned For	Completion
Develop tools to generate authentication and authorization list for accessing gridftp server	IF, REX, FermiCloud	06/27/2011	07/28/2011
Setup a test bed for testing gridftp server	IF, REX, CEDPS	06/29/2011	07/28/2011
Develop condor-globusonline-transfer-plugin	CEDPS, glideinWMS	08/03/2011	08/05/2011
Integrate condor-globusonline-transfer-plugin with glideinWMS	CEDPS, glideinWMS	08/19/2011	08/31/2011
Deploy glideinWMS for end-to-end testing and commissioning	IF, REX, CEDPS	08/26/2011	09/09/2011
Deploy gridftp servers for the required IF experiment(s) *Created base image for future use*	IF, CEDPS, FermiCloud	08/28/2011	09/02/2011
Register the gridftp servers with Globus Online	CEDPS	08/29/2011	09/09/2011
Close the Project	CEDPS	09/30/2011	11/11/2011

9. Project Team

Name	Project Role	Effort (FTE Equivalent)
Burt Holzman	Project Manager	10% (07/01/2011 – 09/30/2011)
Parag Mhashilkar	Project Leader	30% (07/01/2011 – 09/30/2011)
Marko Slyz	Team Member	5% (07/01/2011 – 09/30/2011)
Xi Duan	Summer Student	50% (07/15/2011 – 09/15/2011)

10. Budget and Financial Information

10.1   Personnel Cost

Personal cost is listed in the Project Team section above.

10.2   Hardware Cost

Hardware and the resources utilized for this project are hosted on FermiCloud. Cost estimates depend on the FermiCloud economic model, currently being developed (CD-doc-4399)

11. Outstanding Risks

Risk	Impact	Risk Plan Actions
Support for Globus Online being dropped	Very Low	Solution involving Globus Online has been developed for future operation scenarios and the IF Experiments are not planning to use it in immediate future. The Globus Group at Argonne National Laboratory & University of Chicago maintains Globus Online Services. In case the Globus Online Services project is discontinued, the proposed solution will still work since it depends on the availability of the gridftp server. A new condor file transfer plugin needs to be written or the existing working model needs to be adjusted to use gridftp end point natively.
Delay in migrating the ownership of condor-globusonline-transfer-plugin the to Condor community	Low	glideinWMS project will support the plugin and scripts developed until its ownership is transferred to the Condor community

12. Operations and Support

12.1   Operations

VMs running gridftp services are hosted on FermiCloud resources. Currently, the FermiCloud considers these VMs as production VMs. REX/Experiments should communicate with the FermiCloud group to decide the future support model for existing VMs and for creating additional VMs for other experiments.

12.2   Support

The FermiCloud group supports actual VMs. Developers are available to provide consulting level support in future for questions related to the glideinWMS, gridftp service and the authentication tools deployed on the VMs.

13. Next Steps

There is no current plan to open a new phase of the project. Starting October 2011, Parag Mhashilkar will ramp down the effort level to consulting and emergency maintenance only as required.

During the end-to-end testing, the project team came up with different ideas on modifying the CPN script to use globus-url-copy, which the REX/IF groups can use to tightly integrate the transfers using gridftp servers in their workflows. Stress testing the Globus Online Service revealed some of the shortcomings in the Globus Online and Condor Plugin architecture. These shortcomings are listed on the project home page and have been communicated back to the Globus Online and the Condor projects. Project group encourages the REX/IF groups to evaluate how the Globus Online and the condor file transfer technology evolve before adopting it for production.

Documentation for the project will be maintained at the project home page, as indicated above. Tasks to provide support for the transfer plugin scripts and to follow up with the improvements in the Globus Online and Condor file transfer plugin architecture have been incorporated in the glideinWMS project.

14. Lessons Learned

The following section describes the lessons learned from the project.

14.1 Gridftp Setup on VMs

Gridftp service does not come with a native queuing scheme. This restricts the scale to which a single gridftp service can be used. However, multiple VMs running gridftp service can be deployed behind an LVS service to increase the scalability. Using Globus Online makes this trivial as it supports transfer queues and multiple gridftp servers accessible through a single endpoint.

Creation of the local accounts on the VM interferes with the domain accounts. To avoid conflicts, *ypbind* service should be disabled on the VM hosting gridftp service.

14.2 Globus Online Plugin

Condor plugin architecture is interesting way of integrating the data transfer services with the workload management systems. The scheme can be easily adopted to support arbitrary number of custom file transfer protocols. However, the current Condor plugin architecture is quite restrictive and provides very minimal interface. New version of the plugin architecture to be released in future versions of the Condor is expected to address some of the issues.

While using Globus Online through the transfer plugins, troubleshooting was often complicated. Also, Condor will transfer all the output files to a single destination location including the plugin logs. Retrieving the plugin logs is difficult in this case if the file transfers through plugin itself fails.

Condor does not support multiple transfers with a single plugin invocation. Because of this drawback, plugin cannot take advantage of the bulk transfer mechanism in Globus Online.

Environment available to the plugin is different when invoked for input sandbox transfer than that for output sandbox transfer. This makes it difficult for the plugin to depend on environment variables.

Stress tests revealed that a batch of up to 500 globusconnect connections could cause significant failures (approx. 5% observed). The Globus Online team identified this issue and they plan to provide enhancements to mitigate such issues in future.

During the execution of the project, the group learned several valuable lessons and developed proficiencies in integrating glideinWMS with the data transfer services. The project recommends that the OSG and the Computing Division involve the members of this team in the future related projects.